

64/20

#2



OIPE

ENTERED

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/079,429

DATE: 03/08/2002

TIME: 14:32:43

Input Set : A:\106P3D1 SL Haseltine 021302 Human Genome Sciences.txt

Output Set: N:\CRF3\03082002\J079429.raw

3 <110> APPLICANT: Haseltine et al.
 5 <120> TITLE OF INVENTION: Human DNA Mismatch Repair Proteins
 7 <130> FILE REFERENCE: PF106P3D1
 C--> 9 <140> CURRENT APPLICATION NUMBER: US/10/079,429
 10 <141> CURRENT FILING DATE: 2002-02-22
 12 <150> PRIOR APPLICATION NUMBER: PCT/US95/01035
 13 <151> PRIOR FILING DATE: 1995-01-25
 15 <150> PRIOR APPLICATION NUMBER: 08/468,024
 16 <151> PRIOR FILING DATE: 1995-06-06
 18 <150> PRIOR APPLICATION NUMBER: 08/465,769
 19 <151> PRIOR FILING DATE: 1995-06-06
 21 <150> PRIOR APPLICATION NUMBER: 08/294,312
 22 <151> PRIOR FILING DATE: 1994-08-23
 24 <150> PRIOR APPLICATION NUMBER: 08/210,143
 25 <151> PRIOR FILING DATE: 1994-03-16
 27 <150> PRIOR APPLICATION NUMBER: 08/187,757
 28 <151> PRIOR FILING DATE: 1994-01-27
 30 <160> NUMBER OF SEQ ID NOS: 78
 32 <170> SOFTWARE: PatentIn version 3.0
 34 <210> SEQ ID NO: 1
 35 <211> LENGTH: 2525
 36 <212> TYPE: DNA
 37 <213> ORGANISM: homo sapiens
 39 <220> FEATURE:
 40 <221> NAME/KEY: CDS
 41 <222> LOCATION: (42)..(2312)
 43 <400> SEQUENCE: 1
 44 gttgaacatc tagacgtttc cttggctctt ctggcgccaa a atg tcg ttc gtg gca 56
 45 Met Ser Phe Val Ala
 46 1 5
 48 ggg gtt att cgg cgg ctg gac gag aca gtg gtg aac cgc atc gcg gcg 104
 49 Gly Val Ile Arg Arg Leu Asp Glu Thr Val Val Asn Arg Ile Ala Ala
 50 10 15 20
 52 ggg gaa gtt atc cag cgg cca gct aat gct atc aaa gag atg att gag 152
 53 Gly Glu Val Ile Gln Arg Pro Ala Asn Ala Ile Lys Glu Met Ile Glu
 54 25 30 35
 56 aac tgt tta gat gca aaa tcc aca agt att caa gtg att gtt aaa gag 200
 57 Asn Cys Leu Asp Ala Lys Ser Thr Ser Ile Gln Val Ile Val Lys Glu
 58 40 45 50
 60 gga ggc ctg aag ttg att cag atc caa gac aat ggc acc ggg atc agg 248
 61 Gly Gly Leu Lys Leu Ile Gln Ile Gln Asp Asn Gly Thr Gly Ile Arg
 62 55 60 65
 64 aaa gaa gat ctg gat att gta tgt gaa agg ttc act act agt aaa ctg 296

RAW SEQUENCE LISTING

DATE: 03/08/2002

PATENT APPLICATION: US/10/079,429

TIME: 14:32:43

Input Set : A:\106P3D1 SL Haseltine 021302 Human Genome Sciences.txt

Output Set: N:\CRF3\03082002\J079429.raw

```

65 Lys Glu Asp Leu Asp Ile Val Cys Glu Arg Phe Thr Thr Ser Lys Leu
66 70 75 80 85
68 cag tcc ttt gag gat tta gcc agt att tct acc tat ggc ttt cga ggt 344
69 Gln Ser Phe Glu Asp Leu Ala Ser Ile Ser Thr Tyr Gly Phe Arg Gly
70 90 95 100
72 gag gct ttg gcc agc ata agc cat gtg gct cat gtt act att aca acg 392
73 Glu Ala Leu Ala Ser Ile Ser His Val Ala His Val Thr Ile Thr Thr
74 105 110 115
76 aaa aca gct gat gga aag tgt gca tac aga gca agt tac tca gat gga 440
77 Lys Thr Ala Asp Gly Lys Cys Ala Tyr Arg Ala Ser Tyr Ser Asp Gly
78 120 125 130
80 aaa ctg aaa gcc cct cct aaa cca tgt gct ggc aat caa ggg acc cag 488
81 Lys Leu Lys Ala Pro Pro Lys Pro Cys Ala Gly Asn Gln Gly Thr Gln
82 135 140 145
84 atc acg gtg gag gac ctt ttt tac aac ata gcc acg agg aga aaa gct 536
85 Ile Thr Val Glu Asp Leu Phe Tyr Asn Ile Ala Thr Arg Arg Lys Ala
86 150 155 160 165
88 tta aaa aat cca agt gaa gaa tat ggg aaa att ttg gaa gtt gtt ggc 584
89 Leu Lys Asn Pro Ser Glu Glu Tyr Gly Lys Ile Leu Glu Val Val Gly
90 170 175 180
92 agg tat tca gta cac aat gca ggc att agt ttc tca gtt aaa aaa caa 632
93 Arg Tyr Ser Val His Asn Ala Gly Ile Ser Phe Ser Val Lys Lys Gln
94 185 190 195
96 gga gag aca gta gct gat gtt agg aca cta ccc aat gcc tca acc gtg 680
97 Gly Glu Thr Val Ala Asp Val Arg Thr Leu Pro Asn Ala Ser Thr Val
98 200 205 210
100 gac aat att cgc tcc gtc ttt gga aat gct gtt agt cga gaa ctg ata 728
101 Asp Asn Ile Arg Ser Val Phe Gly Asn Ala Val Ser Arg Glu Leu Ile
102 215 220 225
104 gaa att gga tgt gag gat aaa acc cta gcc ttc aaa atg aat ggt tac 776
105 Glu Ile Gly Cys Glu Asp Lys Thr Leu Ala Phe Lys Met Asn Gly Tyr
106 230 235 240 245
108 ata tcc aat gca aac tac tca gtg aag aag tgc atc ttc tta ctc ttc 824
109 Ile Ser Asn Ala Asn Tyr Ser Val Lys Lys Cys Ile Phe Leu Leu Phe
110 250 255 260
112 atc aac cat cgt ctg gta gaa tca act tcc ttg aga aaa gcc ata gaa 872
113 Ile Asn His Arg Leu Val Glu Ser Thr Ser Leu Arg Lys Ala Ile Glu
114 265 270 275
116 aca gtg tat gca gcc tat ttg ccc aaa aac aca cac cca ttc ctg tac 920
117 Thr Val Tyr Ala Ala Tyr Leu Pro Lys Asn Thr His Pro Phe Leu Tyr
118 280 285 290
120 ctc agt tta gaa atc agt ccc cag aat gtg gat gtt aat gtg cac ccc 968
121 Leu Ser Leu Glu Ile Ser Pro Gln Asn Val Asp Val Asn Val His Pro
122 295 300 305
124 aca aag cat gaa gtt cac ttc ctg cac gag gag agc atc ctg gag cgg 1016
125 Thr Lys His Glu Val His Phe Leu His Glu Glu Ser Ile Leu Glu Arg
126 310 315 320 325
128 gtg cag cag cac atc gag agc aag ctc ctg ggc tcc aat tcc tcc agg 1064
129 Val Gln Gln His Ile Glu Ser Lys Leu Leu Gly Ser Asn Ser Ser Arg

```

RAW SEQUENCE LISTING

DATE: 03/08/2002

PATENT APPLICATION: US/10/079,429

TIME: 14:32:43

Input Set : A:\106P3D1 SL Haseltine 021302 Human Genome Sciences.txt

Output Set: N:\CRF3\03082002\J079429.raw

130		330		335		340		
132	atg tac ttc acc cag act ttg cta cca gga ctt gct ggc ccc tct ggg							1112
133	Met Tyr Phe Thr Gln Thr Leu Leu Pro Gly Leu Ala Gly Pro Ser Gly							
134		345		350		355		
136	gag atg gtt aaa tcc aca aca agt ctg acc tcg tct tct act tct gga							1160
137	Glu Met Val Lys Ser Thr Thr Ser Leu Thr Ser Ser Thr Ser Gly							
138		360		365		370		
140	agt agt gat aag gtc tat gcc cac cag atg gtt cgt aca gat tcc cgg							1208
141	Ser Ser Asp Lys Val Tyr Ala His Gln Met Val Arg Thr Asp Ser Arg							
142		375		380		385		
144	gaa cag aag ctt gat gca ttt ctg cag cct ctg agc aaa ccc ctg tcc							1256
145	Glu Gln Lys Leu Asp Ala Phe Leu Gln Pro Leu Ser Lys Pro Leu Ser							
146	390		395		400		405	
148	agt cag ccc cag gcc att gtc aca gag gat aag aca gat att tct agt							1304
149	Ser Gln Pro Gln Ala Ile Val Thr Glu Asp Lys Thr Asp Ile Ser Ser							
150		410		415		420		
152	ggc agg gct agg cag caa gat gag gag atg ctt gaa ctc cca gcc cct							1352
153	Gly Arg Ala Arg Gln Gln Asp Glu Glu Met Leu Glu Leu Pro Ala Pro							
154		425		430		435		
156	gct gaa gtg gct gcc aaa aat cag agc ttg gag ggg gat aca aca aag							1400
157	Ala Glu Val Ala Ala Lys Asn Gln Ser Leu Glu Gly Asp Thr Thr Lys							
158		440		445		450		
160	ggg act tca gaa atg tca gag aag aga gga cct act tcc agc aac ccc							1448
161	Gly Thr Ser Glu Met Ser Glu Lys Arg Gly Pro Thr Ser Ser Asn Pro							
162		455		460		465		
164	aga aag aga cat cgg gaa gat tct gat gtg gaa atg gtg gaa gat gat							1496
165	Arg Lys Arg His Arg Glu Asp Ser Asp Val Glu Met Val Glu Asp Asp							
166	470		475		480		485	
168	tcc cga aag gaa atg act gca gct tgt acc ccc cgg aga agg atc att							1544
169	Ser Arg Lys Glu Met Thr Ala Ala Cys Thr Pro Arg Arg Arg Ile Ile							
170		490		495		500		
172	aac ctc act agt gtt ttg agt ctc cag gaa gaa att aat gag cag gga							1592
173	Asn Leu Thr Ser Val Leu Ser Leu Gln Glu Glu Ile Asn Glu Gln Gly							
174		505		510		515		
176	cat gag gtt ctc cgg gag atg ttg cat aac cac tcc ttc gtg ggc tgt							1640
177	His Glu Val Leu Arg Glu Met Leu His Asn His Ser Phe Val Gly Cys							
178		520		525		530		
180	gtg aat cct cag tgg gcc ttg gca cag cat caa acc aag tta tac ctt							1688
181	Val Asn Pro Gln Trp Ala Leu Ala Gln His Gln Thr Lys Leu Tyr Leu							
182		535		540		545		
184	ctc aac acc acc aag ctt agt gaa gaa ctg ttc tac cag ata ctc att							1736
185	Leu Asn Thr Thr Lys Leu Ser Glu Glu Leu Phe Tyr Gln Ile Leu Ile							
186	550		555		560		565	
188	tat gat ttt gcc aat ttt ggt gtt ctc agg tta tcg gag cca gca ccg							1784
189	Tyr Asp Phe Ala Asn Phe Gly Val Leu Arg Leu Ser Glu Pro Ala Pro							
190		570		575		580		
192	ctc ttt gac ctt gcc atg ctt gcc tta gat agt cca gag agt ggc tgg							1832
193	Leu Phe Asp Leu Ala Met Leu Ala Leu Asp Ser Pro Glu Ser Gly Trp							
194		585		590		595		

RAW SEQUENCE LISTING

DATE: 03/08/2002

PATENT APPLICATION: US/10/079,429

TIME: 14:32:43

Input Set : A:\106P3D1 SL Haseltine 021302 Human Genome Sciences.txt

Output Set: N:\CRF3\03082002\J079429.raw

```
196 aca gag gaa gat ggt ccc aaa gaa gga ctt gct gaa tac att gtt gag      1880
197 Thr Glu Glu Asp Gly Pro Lys Glu Gly Leu Ala Glu Tyr Ile Val Glu
198      600      605      610
200 ttt ctg aag aag aag gct gag atg ctt gca gac tat ttc tct ttg gaa      1928
201 Phe Leu Lys Lys Lys Ala Glu Met Leu Ala Asp Tyr Phe Ser Leu Glu
202      615      620      625
204 att gat gag gaa ggg aac ctg att gga tta ccc ctt ctg att gac aac      1976
205 Ile Asp Glu Glu Gly Asn Leu Ile Gly Leu Pro Leu Leu Ile Asp Asn
206 630      635      640      645
208 tat gtg ccc cct ttg gag gga ctg cct atc ttc att ctt cga cta gcc      2024
209 Tyr Val Pro Pro Leu Glu Gly Leu Pro Ile Phe Ile Leu Arg Leu Ala
210      650      655      660
212 act gag gtg aat tgg gac gaa gaa aag gaa tgt ttt gaa agc ctc agt      2072
213 Thr Glu Val Asn Trp Asp Glu Glu Lys Glu Cys Phe Glu Ser Leu Ser
214      665      670      675
216 aaa gaa tgc gct atg ttc tat tcc atc cgg aag cag tac ata tct gag      2120
217 Lys Glu Cys Ala Met Phe Tyr Ser Ile Arg Lys Gln Tyr Ile Ser Glu
218      680      685      690
220 gag tcg acc ctc tca ggc cag cag agt gaa gtg cct ggc tcc att cca      2168
221 Glu Ser Thr Leu Ser Gly Gln Gln Ser Glu Val Pro Gly Ser Ile Pro
222      695      700      705
224 aac tcc tgg aag tgg act gtg gaa cac att gtc tat aaa gcc ttg cgc      2216
225 Asn Ser Trp Lys Trp Thr Val Glu His Ile Val Tyr Lys Ala Leu Arg
226 710      715      720      725
228 tca cac att ctg cct cct aaa cat ttc aca gaa gat gga aat atc ctg      2264
229 Ser His Ile Leu Pro Pro Lys His Phe Thr Glu Asp Gly Asn Ile Leu
230      730      735      740
232 cag ctt gct aac ctg cct gat cta tac aaa gtc ttt gag agg tgt taa      2312
233 Gln Leu Ala Asn Leu Pro Asp Leu Tyr Lys Val Phe Glu Arg Cys
234      745      750      755
236 atatggttat ttatgcactg tgggatgtgt tcttctttct ctgtattccg atacaaagtg      2372
238 ttgtatcaaa gtgtgatata caaagtgtac caacataagt gttggttagca cttagactt      2432
240 atacttgctt tctgatagta ttcttttata cacagtggat tgattataaa taaatagatg      2492
242 tgtcttaaca taaaaaaaaa aaaaaaaaaa aaa      2525
245 <210> SEQ ID NO: 2
246 <211> LENGTH: 756
247 <212> TYPE: PRT
248 <213> ORGANISM: homo sapiens
250 <400> SEQUENCE: 2
252 Met Ser Phe Val Ala Gly Val Ile Arg Arg Leu Asp Glu Thr Val Val
253 1      5      10      15
256 Asn Arg Ile Ala Ala Gly Glu Val Ile Gln Arg Pro Ala Asn Ala Ile
257      20      25      30
260 Lys Glu Met Ile Glu Asn Cys Leu Asp Ala Lys Ser Thr Ser Ile Gln
261      35      40      45
264 Val Ile Val Lys Glu Gly Gly Leu Lys Leu Ile Gln Ile Gln Asp Asn
265      50      55      60
268 Gly Thr Gly Ile Arg Lys Glu Asp Leu Asp Ile Val Cys Glu Arg Phe
269 65      70      75      80
```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/079,429

DATE: 03/08/2002

TIME: 14:32:43

Input Set : A:\106P3D1 SL Haseltine 021302 Human Genome Sciences.txt

Output Set: N:\CRF3\03082002\J079429.raw

```
272 Thr Thr Ser Lys Leu Gln Ser Phe Glu Asp Leu Ala Ser Ile Ser Thr
273      85                      90                      95
276 Tyr Gly Phe Arg Gly Glu Ala Leu Ala Ser Ile Ser His Val Ala His
277      100                    105                    110
280 Val Thr Ile Thr Thr Lys Thr Ala Asp Gly Lys Cys Ala Tyr Arg Ala
281      115                    120                    125
284 Ser Tyr Ser Asp Gly Lys Leu Lys Ala Pro Pro Lys Pro Cys Ala Gly
285      130                    135                    140
288 Asn Gln Gly Thr Gln Ile Thr Val Glu Asp Leu Phe Tyr Asn Ile Ala
289 145      150                    155                    160
292 Thr Arg Arg Lys Ala Leu Lys Asn Pro Ser Glu Glu Tyr Gly Lys Ile
293      165                    170                    175
296 Leu Glu Val Val Gly Arg Tyr Ser Val His Asn Ala Gly Ile Ser Phe
297      180                    185                    190
300 Ser Val Lys Lys Gln Gly Glu Thr Val Ala Asp Val Arg Thr Leu Pro
301      195                    200                    205
304 Asn Ala Ser Thr Val Asp Asn Ile Arg Ser Val Phe Gly Asn Ala Val
305      210                    215                    220
308 Ser Arg Glu Leu Ile Glu Ile Gly Cys Glu Asp Lys Thr Leu Ala Phe
309 225      230                    235                    240
312 Lys Met Asn Gly Tyr Ile Ser Asn Ala Asn Tyr Ser Val Lys Lys Cys
313      245                    250                    255
316 Ile Phe Leu Leu Phe Ile Asn His Arg Leu Val Glu Ser Thr Ser Leu
317      260                    265                    270
320 Arg Lys Ala Ile Glu Thr Val Tyr Ala Ala Tyr Leu Pro Lys Asn Thr
321      275                    280                    285
324 His Pro Phe Leu Tyr Leu Ser Leu Glu Ile Ser Pro Gln Asn Val Asp
325      290                    295                    300
328 Val Asn Val His Pro Thr Lys His Glu Val His Phe Leu His Glu Glu
329 305      310                    315                    320
332 Ser Ile Leu Glu Arg Val Gln Gln His Ile Glu Ser Lys Leu Leu Gly
333      325                    330                    335
336 Ser Asn Ser Ser Arg Met Tyr Phe Thr Gln Thr Leu Leu Pro Gly Leu
337      340                    345                    350
340 Ala Gly Pro Ser Gly Glu Met Val Lys Ser Thr Thr Ser Leu Thr Ser
341      355                    360                    365
344 Ser Ser Thr Ser Gly Ser Ser Asp Lys Val Tyr Ala His Gln Met Val
345      370                    375                    380
348 Arg Thr Asp Ser Arg Glu Gln Lys Leu Asp Ala Phe Leu Gln Pro Leu
349 385      390                    395                    400
352 Ser Lys Pro Leu Ser Ser Gln Pro Gln Ala Ile Val Thr Glu Asp Lys
353      405                    410                    415
356 Thr Asp Ile Ser Ser Gly Arg Ala Arg Gln Gln Asp Glu Glu Met Leu
357      420                    425                    430
360 Glu Leu Pro Ala Pro Ala Glu Val Ala Ala Lys Asn Gln Ser Leu Glu
361      435                    440                    445
364 Gly Asp Thr Thr Lys Gly Thr Ser Glu Met Ser Glu Lys Arg Gly Pro
365      450                    455                    460
368 Thr Ser Ser Asn Pro Arg Lys Arg His Arg Glu Asp Ser Asp Val Glu
```

Use of n and / or Xaa has been detected in the
Sequence Listing. Review the Sequence Listing
to ensure a corresponding explanation is present
in the <220> to <223> fields of each sequence
using n or Xaa.

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/079,429

DATE: 03/08/2002

TIME: 14:32:44

Input Set : A:\106P3D1 SL Haseltine 021302 Human Genome Sciences.txt

Output Set: N:\CRF3\03082002\J079429.raw

L:9 M:270 C: Current Application Number differs, Replaced Current Application Number

L:1537 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:17